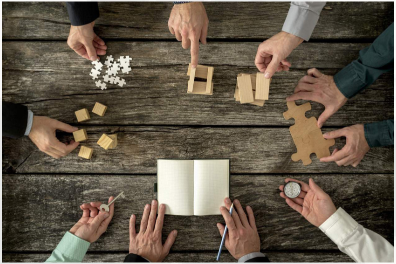
An Epistemic Logic for Modelling Cooperative Agents

Stefania Costantini[1] Andrea Formisano[2] Valentina Pitoni[1]

DISIM - Università degli Studi di L'Aquila[1]
DMIF - Università degli Studi di Udine[2]

VardiFest 2022

Dynamic Logic of INFerable



We discuss features of our epistemic logic framework *L-DINF*.

The focus is on modeling the group dynamics of cooperative agents, aimed to reach common objectives. Cooperation may concern various aspects, e.g., sharing the *cost* of actions, and respecting group members' preference and roles about performing actions. We also model group dynamics involving aspects of the "Theory of Mind", i.e., the ability to reason about the beliefs of others.

Characterization

L-DINF (which owes to past work by Balbiani, Duque and Lorini) is aimed to model intelligent agents, that interact with an external environment and so, beyond their background knowledge, have beliefs, acquire perceptions and develop new beliefs.

- The *static component*, L-INF (used to represent the“Long-Term Memory of the agent”), is a logic of explicit beliefs and background knowledge (the latter satisfying the “omniscience” principles).
- The *dynamic component*, L-DINF (used to represent the “Short-Term Memory of the agent”), encompasses dynamic inferential operations on perceptions, producing new beliefs.

Inferential (Mental) Operations

Inferential Operations are the basic ways for an agent of forming new explicit beliefs via inference:

- $\cap(\varphi,\psi)$ is the inferential operation which consists in closing the explicit belief that φ and the explicit belief that ψ under conjunction;
- $\downarrow(\varphi, \psi)$ consists in inferring ψ from φ in case φ is believed and,according to an agent's background knowledge, ψ is a logical consequence of φ;
- $\vdash(\varphi,\psi)$ is the inferential operation which consists in inferring ψ from φ in case φ is believed and, according to the agent's working memory, ψ is a logical consequence of φ;
- $\dashv(\varphi, \psi)$: this inferential action performs a simple form of "belief revision".

Mental Operations and Physical Actions

- $[G : \alpha]\varphi$ should be read "φ holds after the inferential action α has been performed by at least one of the agents in G, and all agents in G have common knowledge about this fact".

The formula φ which is inferred and asserted as a new belief by an inferential action can be $can\_do_i(\phi_A)$ or $do_i(\phi_A)$, which denotes the actual [possibility of] execution of physical action ϕ_A.

The do_i^P beliefs constitute a *history* of the agent's operation.

Models of $L\text{–}DINF$

A *L-INF model* is a tuple $M = (W, N, \mathcal{R}, E, B, C, A, H, P, V)$ where:

W is a set of worlds (or situations);

$\mathcal{R} = \{R_i\}_{i \in Agt}$ is a collection of equivalence relations on W
it models the background knowledge of each agent

$N : Agt \times W \longrightarrow 2^{2^W}$ is a neighborhood function such that

- if $X \in N(i, w)$ then $X \subseteq \{v \in W \mid wR_iv\}$
- if wR_iv then $N(i, w) = N(i, v)$

it models the epistemic state of each agent

Models for $L\text{–}DINF$ (cont'd)

A *L-INF model* is a tuple $M = (W, N, \mathcal{R}, E, B, C, A, H, P, V)$ where:

$E : Agt \times W \longrightarrow 2^{\mathcal{L}_{\mathsf{ACT}}}$ is an executability function of mental actions

$B : Agt \times W \longrightarrow \mathbb{N}$ is a budget function

$C : Agt \times \mathcal{L}_{\mathsf{ACT}} \times W \longrightarrow \mathbb{N}$ is a cost function for mental actions

$A : Agt \times W \longrightarrow 2^{Atm_A}$ is an executability function for physical actions

$H : Agt \times W \longrightarrow 2^{Atm_A}$ is an enabling function for physical actions

$P : Agt \times W \times Atm_A \longrightarrow \mathbb{N}$ is a preference function for physical actions

$V : W \longrightarrow 2^{Atm}$ is a valuation function

Truth of formulas is determined as one expects, w.r.t. a model M and a world w, and mental actions determine neighbourhood update.

Example: Joint Goals and Roles

Problem: Consider a group of four agents, who are the crew of an ambulance, including a driver, two nurses, and a medical doctor. The driver is the only one enabled to drive the ambulance. The nurses are enabled to perform a number of tasks, such as, e.g., administer a pain reliever, or clean, disinfect and bandage a wound, measure vital signs. It is however the task of a doctor to make a diagnosis, to prescribe medications, to order, perform, and interpret diagnostic tests, and to perform complex medical procedures.

Upon an emergency call, the group will reason, set the intention/goal

$$\mathbf{K}_i(intend_G(rescue\_patient))$$

and devise a plan to achieve it.

For each physical action ϕ_A required by the plan, there will be some agent (let us assume for simplicity only one), for which $do_i(\phi_A))$ will be concluded. In our case:

- For the agent ‘driver’, the group will conclude: $do_{driver}(drive\_to\_patient))$ and $do_{driver}(drive\_to\_hospital))$.
- For the agent ‘doctor’, the group will conclude, e.g., $do_{doctor}(stabilize\_patient))$ and $do_{doctor}(pneumothorax\_aspiration)$.

An interesting point concerns derogation, i.e., for instance, life or death situations where, unfortunately, no-one who is enabled to perform some urgently needed action is available; in such situations perhaps, anyone who is capable to perform this action might perform it.

Enhanced Axiomatization

To allow roles to be dynamically interchanged when needed, axiomatization of actions enabling and execution can be formulated as follows.

(t4') $M, w \models able\_do_i(\phi_A)$ iff $\phi_A \in A(i, w)$

(t4'') $M, w \models enabled\_do_i(\phi_A)$ iff $\phi_A \in A(i, w) \cap H(i, w)$

(t4''') $M, w \models can\_do_i(\phi_A)$ iff $enabled\_do_i(\phi_A) \vee$
$(able\_do_i(\phi_A) \wedge \phi_A \in \mathbf{S}(\mathbf{i}, \phi_\mathbf{A}, \mathbf{w}) \wedge \nexists j \in G : enabled\_do_j(\phi_A))$

(t5') $M, w \models can\_do_G(\phi_A)$ iff $(\exists i \in G : can\_do_i(\phi_A))$

Example: the “Sally-Anne” Task

We formalized the “Sally-Anne” task, which was (in relevant experiments) administered to children to test their ability to develop a “Theory of Mind” concerning the expectation of how someone will act based on the awareness of that person’s false beliefs.

Task specification

1. a child (or an agent) is told a story about two girls, Sally and Anne, who are in a room with a basket and a box.
2. Sally puts a doll into the basket, then leaves the room, and, in her absence, Anne moves the doll to the box.
3. The child is then asked: “where does Sally believe the doll to be?”.

To pass the test, the child should answer that Sally believes the doll to be in the basket.

Example: the "Sally-Anne" Task in L-DINF

Assume to have a group involving agents 1 (Sally), 2 (Anne), and 3 (Observer); initially, they share their beliefs and, when Sally puts the doll in the box, we will easily have that $\mathbf{B}_i(in(doll, box))$ for $i = 1, 2, 3$.

Each agent i of the group is aware, at a certain point, that Sally (agent 1) left the group.

If Anne moves the doll into the basket, only agents $i = 2, 3$ are able to conclude that $\mathbf{B}_i(in(doll, basket))$.

Reasoning in someone else's shoes

We introduced modality $\mathrm{B}_{i,j}$, applied here to $\mathrm{B}_{3,1}$, that makes agent 3 able to "impersonate" agent 1, i.e., to reason within agent 1's beliefs, or, precisely, within the last version of agent 1's beliefs owned by agent 3.

So, given question: $\mathbf{B}_{3,1}(in(doll, Place))$? Agent 3 concludes immediately

$$\mathbf{B}_{3,1}(in(doll, box)).$$

We have seen that agents 2 and 3 have instead concluded $in(doll, basket)$, inference that agent 3 cannot do when she puts herself "in the shoes" of agent 1, which, having left the group, does not have the necessary beliefs available, which were formed after she left.

Conclusions

- We have shown, also by means of examples, features of our logic *L-DINF*.
- Formal results: we have shown that *L-DINF* is: **sound**, using a particular axiomatization; **strongly complete**, using a canonical model argument.
- We intend in **future work** to extend this logic, including explicit time and evolution of roles in time under changing circumstances, and more involved interactions among groups of agents.